(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79449 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/08656

(22) International Filing Date: 16 April 2001 (16.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/552,929 18 April 2000 (18.04.2000) US
09/770,160 26 January 2001 (26.01.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 09/552,929 (CIP)
Filed on 18 April 2000 (18.04.2000)
US 09/770,160 (CIP)
Filed on 26 January 2001 (26.01.2001)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., TOM** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US). **DRMANAC, Radoje, T.** [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsksy and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/79449 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

SEQ ID NO: of nucleo-tide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USSN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
						RKAASQSDKPAEKKEDESQMEDPS TSPSPGTRAASQPPNSSK\AGRKPW DRNNPLRNPLSSNLVRNPLLAKGPR KLRAPFSQQPHSRMKPAGS\VSDMA \LDAFDLD\RMKQEI*KEVVRELHK GERKEIID\AIRQEA*SGISRKKNLGH RAHPPTRTSFICSQRPRLM
1746	7243	A	1876	1	668	GERGVARHDRPRGTLREYKVVGRC LPTPK\CHTPPL\YR\MRIFAP*SMSSL SPRF\WYFVSQLKKDEESLQWRFSY CAQVFEKSP\LRVK\NFGIWLR\YDS RSG\THNMY\REY\RDLDHPQAPVHP SCLTRDNGVAPAPAA/HEAHFHFRFI ERLEEI\AGQQDCRRPGCSKQFPRIS RFKFPAAPPGSLRRQDKPRF\TTKRP KTFLKVQGPSSGVCPQNKTQETPR
1747	7244	A	1877	1	1059	
1748	7245	A	1878	87	260	
1749	7246	A	1879	1	1254	
1750	7247	A	1880	160	615	PSLNTYVTSPLSENFSARYRNHSND LTCVHTELQNKTKLTVLEGDILDEP FLKRACQ\DVSVI\IHTACIIDVFGV\T HRESIMNVNVKGRVAWGGDKARW GNEDQKEGQEGKRSLSIEHLLCSGP SDFADHYQLGELKAAIFSFIDEKTRT EQ
1751	7248	A	1881	53	1338	CPLQGHPRVTLESDLLPSIFCFLVSD SCYFGLATMGWSCLVTGAGGLLGQ RIVRLLVEEKELKEIRALDKAFRPEL REEFSKLQ\NK\TKLTVLEGDILDEPF LKESLARDRLRSIIHTACFHLMSFGV \THREFF\MNVQC*KVPSSC*EACVQ ASVPVFIYTSSIEVAGPNSYKEIIQNG HEEEPLEN\TWPAPYPRSKKLA\KKA VLAANGWNLK\NGGALYTCALRPM YIYGEGSRFLSVSINEALNNNGILSS VGKFST\VNPVYV\GNVAWGHILAL RALQDPKKAPSIRGQFYYISDDTPH QSYDNLNYTL\SKE\FGPPPLDSRW\S FPLSLMYW\IGFLLGNR*GFLL\RPIY TYRPPFNRHISSHCSN*ALFHLLFIKE GFSEILGVLRPLLTAGGGKAKAGKR VGSWVWVPFVDPAQGRNLEVPRIQ
1752	7249	A	1882	3	575	HSLFGTSEVINKLLVPDA\MGHFTEE D\KATI\TSLWGK\VNVE\DAGGE\TP GKGSLVVYP\WTQRF\FD\SFGNLSS ASAI\MGKPPKSKAHG\KKVLTFLGT MPTKHLE*FSRGTFCPSLK*TCTC*Q ACMWDPGGTFKLPGENVAGLTVFG QSHFRQKNFTPEGARFFLGRKMGD LELASALVPSRLPLKPLGP
1753	7250	A	1883	1	960	GRPAPEDGGPLSLPNAAMARGPKK HLKRVAA\PKHWMLDKLTGVFAPR PSTGPHKL\RECLPFIIF\LRNRLKYA LTS\DEVKKICMQRFIKI\DGQVR\TD ITYP\AGFMDVI\SIDKDGREFSVL/Y LIDTQGVRFCL*HRITP*GRAKVQSC AKMRKILLWAPKGIPSSWVT\HDAR

SEQ ID NO: of nucleo-tide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USSN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
						NHPATPDPPSSKVN*YHFRLDLETG KDYLISSKFDTW*PCVMVT\GGA\N LGRNWVLITN\RERHPGIF*PLVHVK \DANGNKLLATSDFSNIFWLLGKGN KPW\ISL\PRGKGIPPHHLLEERDKRL AAKQSSWVKWGPWVTWSDLLVP
1754	7251	A	1884	1	1218	FFQNSARGAGAGWQLPWTRFVWT\ SGLLEINE\TLVIQQRGVRIYDGEEKI KFDAGTLLLSTHRLIWRDQKNHEC CMAILLSQI\VFIEEQA\AGIGKSAKI VVHL\HPAPPNKEPGP\FQSSKNSYI KLSFKEHGQIEFYRRLSEEMTQRRW ENMPVSQSLQTNRGPQPGRIRAVGI VGTERKLEEKRKETDKNISEAFEDL SKLMIKAKEMVELSKSIANKIKDKQ GDITEDETIRFKSYL\LSMGIANPVT RETYGSGTQYHM\QLAKQL\AWNIA RVPLEERGGIMSLTEVYCLVNRARG MELLSPEDLVNACKMLEALKLPLR LRVFDSGVMVIELQSHKEEEMVAS ALETVSEMGSLTS*EFAKLVGMSVL LAKERLLLAEKMGHLCRDDSVEGL RFYPNLFMTQS
1755	7252	C	1885	179	361	MPKVCFVHNFLKTSSERDLFALMN TVGKKHSIMSEKGRSKKFLHLIDSK KNEDPHLDGTL*
1756	7253	A	1886	2	913	RRLLLFGWARSGAVSLGSAGVSSS GFLTAPHSRRLTAAAAAAGGAWRF EAERHRGWGAEEEQQPEGGAVCPG TERPCAMAYAYLFKYIIIGDTGGGR\ SCLLLQFTDKRFQPSAMTLTNGVEF GARMITIDGKQIK\LQIW\DTAGQES\ FRSITR\SYY\RGAAGALLVYDITR\R DTSTHLTTW\LEDA\RQHSHFQHGS LCLLGNKSDL\ESRKE/VSKKRKEGE SFLQPRNHGLHLPWKTSCKNCFPM* KEAFINTSKRNFIEKIQ\EGVFDINNE A\NGIKIGP\QHAATNATHAG\NQGG QQAGGGCC
1757	7254	A	1893	138	426	FIHSHCCIVFRLFIHFSLHPKVIHSPIN SLLRIFQF*AIMNSTV*NILIHVFW*V YTFPF\GINPKKGIARL*GVYIFSFSIY CQTVFQSDCKKAPF
1758	7255	A	1894	45	1057	FLVFLVETGFHHVAQAVLELLASSD PPALAPPKCWDYRCELLRLAEFCFL RTEFWYLLFFFFWRRSLALSPRLEC SGANL\THCNLR/LPGFKQFSCLSLSS SWDYRCMPPHLATFFVF/SVETGFH RVAQASLELLSSGSLPALA/FPKC\W DYRAKATV/WPSPGVSSFILGL*TS* FHSLEPYLHAWKTTSHLPTKEALT W/VSHTAKTKHLWILVSILMEF*VA LIS/SFFLGPGGK*T*VTAPQCPSLGQ DTLS*FLHAACTRSVPYPGLA/CGPS LWLTRVLLLPTPP*QQHNP/DTLEKT SFPGPHWIL*/TPQPSLSETPAPKVPP FPAFGSIPTHEEPGLP
1759	7256	A	1895	2	289	

SEQ ID NO: of nucleo-tide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USSN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
1760	7257	A	1896	1	397	
1761	7258	A	1897	1	410	STMISPVLILFSSFLCHVAIAGRTCPK PDDLPFSTVVPLKTFYEPGEEITYSC KPGYVSRGGM\RKFICPLTGLWPIN TLKCTPRVCPFAGNLRKMGAVRLIT DFLNYSPTRFSFSLLTWGFILEWAL DS\AKCIEGG
1762	7259	A	1898	19	1215	CQCDSSTMIFSRCSSLFSSFLCHVAI AGRTCPKPDDLPFSTVVPLKTFYEP G\EEITYSCKPGYVSRGGIEESLSCPL \TGTVGPFNTSGNVTPRVCPF\AGIFR KMGGRTLITTF*NYPNTDPVFSLLTL GF*FWNGALDFWPSCTGGKGKW\S P\ELPGLVAPII\CPP\PSIP/TGFATLH VLLRPFRLGNNSPPIGDTAVFECLA HNMAMFG\NDTIT\CTTHGKLDLNY PECRGSKMPPFPHQDPDNGIW*TYP CQNPNTLFTRVKAPHLGLPHDGIFS GMGPRKE\EC*PQTWGKPGSWPLA PSW*KPSLVKGTPVKKRPTVV/YPQ GERVKDSREKFKEWECLHG**KFLS FCKNKEKKCSYTEDAQCIDGTIEVP KCFK\EHSSLAFWKT\DAS\DVKPC
1763	7260	A	1899	58	446	
1764	7261	A	1900	1	954	MGEVSGTSDCTDDQCRQVKKALEG GKAARGHRSKIKIRFFRPGGLGPGP AITAVAGMPRVYIGRLSYQAREHA V\ERLLNGHAKILEVDLKNGYGFVE FDDLRDADDAVYELNGKDLCGERV IVEHARGPRRDGSYGSGRSGYGYR RSGRDKYGPPTRTEDRLIVEN\LTSR CSWQDLKDYMRQAGEVTYADAHK GRQKMKGVIEFVSYSDMKRALEKL DGTEVNGRKIRLVEDKPGSRRRRSY SRSR\SHSRSRSRSRHSRKSRSRSGSS KSSHSKSRSRSRSGSRSRSKSRSRSQ SRSRSKKEKSRSPSKDKS\RSRSHSA\ GKSRSKSKDQAE\EKFQNNDNV\GK PKSRSPSRHKSKSKSRSRSQERRVEE GRKRGSF*QGQ/EAQEKSLRQSRSR\ SRSKAGSR*PVDRSRSKSKDKRKSR KRSREESRSRSRSRSKSERSRKRG\S KRDSKAS\SCKKKKKEDTDRSQSRS PSRSV\SKEREHA/RSLESSQREGRG ESENAGTNQEDPGPGPRSN\SKSKP NLPIRMHRSKIKSQASKTPISGPMSR SR\SASRSP\SRSRSKSRSRSQSRSRS KKEKSRSPSKDKSLQPQP
1765	7262	A	1901	3	180	
1766	7263	A	1902	227	440	GMHNVCYVAVNE*FCGFIIR*SLAE RRQIS*EFQLFKFTLCLELILARRAC RESMASPVAGSWSHFPEREF
1767	7264	A	1903	2	438	HEELDTSERKIEFDSASGTYTLYLI\I GDAHFEEPQSLWNVADLVHQSPPE EKAPLDLSCPQNLFTPK\QEIQWIRI GA\NVS\NFTFAP\STIIFH\LGHA\AM LGLMYVYWTQLNMF\QTLKYLAIL GSVTFLAGNRMLAQQAVKRTAH

SEQ ID NO: of nucleo-tide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USSN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
1768	7265	A	1904	1	1660	
1769	7266	A	1905	156	2369	PVLKTHPGPQSLPRVPGVPCGGLLE PLSRAEVSPRFGLRRDLLGGMAPPG SSTVFLLALTIIASTWALTPTHYLTK HDVERLKASLDRPFTNLESAFYSIV GLSSLGAQVPDAKKACTYIRSNLDP SNVDSLFYA\AQA\SQGLSGCEISISN ETKDLLLA\AVSE\DSSVYPRSYHAS WQL*SGLLGLSLWAVPKESTQVAL NWLVFKQGKETVL\ATVQALQTAS HLSQQADLRSIVEEIEDLVARLDEL GGLYLQ\FEEGLETTAL\FVAATYKA /LMDH\VGTE\PSIKE\DQVIQLMNAI F\SKKNFES\LSEAFSV\ASG\AAVLS HNRYHVPVVVVPEGSASDTHEQAI LRLQVTNVLSQPLTQATVKLEHAK SVASRATVLQKTSFTP\VGIVFELNF MNVKFSGG*CDF\LVEVEGDNRYIS\ NTVELRVQDPPTEVGITNVDLSTV\ DKDQSIAP\QTTRVTYPAKAKGTFH SAGQATRNFGLVLSSW*DVNTG\AE LTPHQTFVRLHNQKTGPGSGCLFAE PGQQGTCYKFELDTSERKGLNLTSR SGTYTLYLIIG*CQL*RTQILWKCGL MWVI\KFP*GKEASFDCLCSQEPFSL PKQGNFRHLFPGRP*GRRAPPPWCP NTFTAPESFFGPLL/LCFLRLLWIRD WVPKCLPTFTFCFLSTIIFHPWDML AYAGTSMYVY*TQAQPCSQTLEVP WPILGQCDRFLAGQSGMLAPARQV KRIAAEQSSRLAKYRTLRTAH
1770	7267	A	1906	37	404	PQLSRCRSECMYVNPTVVMTSMGQ ATWSDPHKAKTMLNRIPLGKFAGE SGGSPASVVPAVPVCALGRGGRER WAAASFLYAPDPRPAH\EVEHVVN AILFLLSDRSGMTTGSTLPVEGGFW AC
1771	7268	A	1907	271	1086	YTQCPGIEPVCVDLGDWEATERAL GSVGPVDLLVNNAAVALLQPFLEV TKEAFDR*ACEGGGTSGRGCPGGRS SPNL*PGSVPRPLDPLRVNLRAVIQV SQIVA\RGLI\ARGVPTGPS*NVSSQC FPAGQ*TNHSVLLLPTKGVPLDMLD QG*WAL\ELGPHKLSRCRSGVNA\V NPHSGG*RSMGPGPPWSDPHK\AKI MLNRIP\LGKFAGESEVEHVVNA\IL FLLSDRSGMTTGS\TLPVEGGFWAW LSSLHTPQAPWACFILTPNPSNKT
1772	7269	A	1908	2	305	ARESGSLVAPRSRPPWEHGLPGEHS *DAPRPHKSPTLPWLPHLHLSKEAL DTHQRSQHE\ECMPLYKFTPTSEKR PQLMLPLPEQQCEQLCRFGSTPVTW A
1773	7270	A	1909	2	529	GTVAACGACYWLLGLMAVRASFE NNCEIGCFAKLTNTYCLVAIGGSEN FYSVFEGELSDTIPVVHASIAGCRIIG RMCVG\TEEILADVLKVEVFRQTVA DQVLVGSYCVFSNQGGLVHPKTSIE

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-5497, a mature protein coding portion of SEQ ID NO: 1-5497, an active domain of SEQ ID NO: 1-5497, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-5497.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-5497, a mature protein coding portion of SEQ ID NO: 1-5497, an active domain of SEQ ID NO: 1-5497, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-5497, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 5498-10994, the mature protein portion thereof, or the active domain thereof.

21 The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprises the sequence information of at least one of SEQ ID NO: 1-5497.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

28 A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.